APPENDIX G

Classification and Regression Tree (CART) Analysis for Chicago-Naperville, IL-IN-WI, 2008 8-Hour Ozone Nonattainment Area, 2000-2014

This page left intentionally blank.

Classification and Regression Tree (CART) Analysis for Chicago-Naperville, IL-IN-WI, 2008 8-Hour Ozone Nonattainment Area

A classification and regression tree (CART) analysis is a statistical tool to classify data. Here, it is applied to 8-hour ozone and meteorological data to determine the meteorological conditions most commonly associated with high-ozone days. Once days are classified by their meteorology, ozone concentration trends among days with the same conditions can be developed. By examining trends only on days with similar meteorology, the influence of year-to-year meteorological variability on ozone concentrations is minimized and we assume that any remaining trend is the result of reductions in emissions of ozone precursors and other non-meteorological factors.

This CART analysis was conducted by LADCO using 8-hour ozone monitoring data for two regions of Chicago-area ozone sites. The first analysis area is comprised of two northern monitors: Chiwaukee (ID number 55-059-0019) and Zion (17-097-1007). The second analysis area is represented by a combination of multiple southern monitoring sites in Cook County, Illinois including: Alsip (ID number 17-031-0001), Chicago-SWFP (17-031-0032), Chicago-Taft (17-031-1003), Lemont (17-031-1601), Cicero (17-031-4002), Northbrook (17-031-4201), and Evanston (17-031-7002). The analysis included data from the years 2000-2014, which encompasses many years prior to the promulgation of the 2008 ozone NAAQS. This analysis therefore addresses long-term trends rather than the direct impacts of the 2008 ozone NAAQS. The goal of the analysis was to determine the meteorological conditions associated with high ozone episodes in the Chicago air-shed and to construct trends for the days identified as sharing similar meteorological characteristics.

The CART analyses for the Chicago-area ozone study processed multiple meteorological variables for each day to determine which are the most effective at predicting ozone concentrations. Meteorological data collected for the Chiwaukee/Zion monitors were taken from Mitchell Field (Milwaukee) NWS station and processed by LADCO. Upper air observations were taken from the Green Bay, Wisconsin NWS site. Meteorological data collected for the analysis of southern monitors was taken from the Chicago O'Hare Airport National Weather Service (NWS) station and processed by LADCO. Upper air observations, taken from Lincoln, Illinois NWS site, were downloaded from the National Climatic Data's Center (NCDC) Integrated Global Radiosonde Archive. Meteorological variables for both analyses included maximum and average daily temperatures, dew points, relative humidity and air pressure at the surface and different levels of the atmosphere, wind directions and wind speeds, change in temperatures and air pressure from the previous day, average wind speeds and temperatures over a 2 or 3-day period, day of the week, cloud cover, daily precipitation, and many other parameters.[1]

[1] The original meteorological database used to support this effort, called MetDat, was developed by EPA Office of Air Quality Planning and Standards (OAQPS) and subsequently revised by both Sonoma Technology and LADCO.

Regression trees, where each branch describes the meteorological conditions associated with different ozone concentrations, were developed to classify each summer day (May – September). Although the exact selection of predictive variables changes from site to site, the universally common predictors are temperature, wind direction, and relative humidity. These are included in the dataset as daily averages and maximums as well as averages at specific times throughout the day (morning 7-10 am, afternoon 1-4 pm, etc.). Similar days were assigned to nodes, which are equivalent to branches of the regression tree. By grouping days with similar meteorology, the influence of meteorological variability on the underlying trend in ozone concentrations is partially removed; the remaining trend is presumed to be due to trends in precursor emissions or other non-meteorological influences. Ozone trends in these nodes were then plotted.

The CART analysis determined that there were four nodes from the Chiwaukee Prairie and Zion monitors and two nodes from the Cook County monitors that demonstrated similar meteorological conditions and had the strongest correlation with high ozone episodes (greater than 55 parts per billion (ppb) of ozone). Tables 1 and 2 categorize each analysis and the shared meteorological conditions for each high-ozone node along with the frequency and average ozone concentrations. All of the high-ozone nodes had high maximum temperatures, and many were distinguished by southerly or southeasterly winds and/or low relative humidity.

Table 1: Meteorological conditions, occurrence, and average ozone for the four nodes identified with high ozone concentrations for the Chiwaukee Prairie and Zion monitors.

<table>
<tr><th rowspan="2">Conditions</th><th colspan="4">Node</th></tr>
<tr><th>16</th><th>19</th><th>20</th><th>21</th></tr>
<tr><td>Maximum temperature</td><td colspan="4">> 77.56 °F</td></tr>
<tr><td>Average morning temperature</td><td>≤ 77.56 °F</td><td colspan="3">> 77.56 °F</td></tr>
<tr><td>Southerly component of 24-hour transport distance</td><td>> -131.21 km</td><td colspan="3" rowspan="3"></td></tr>
<tr><td>Average relative humidity</td><td>≤ 77.09%</td></tr>
<tr><td>Deviation in morning height of 700 mb surface from 10-year monthly mean</td><td>> -10.75 m</td></tr>
<tr><td>Southerly component of average afternoon winds</td><td rowspan="3"></td><td>≤ 2.41 m/s</td><td colspan="2">>2.41 m/s</td></tr>
<tr><td>Maximum temperature</td><td>> 85 °F</td><td colspan="2"></td></tr>
<tr><td>Easterly component of average afternoon winds</td><td></td><td>≤ 1.88 m/s</td><td>> 1.88 m/s</td></tr>
<tr><td>Average ozone</td><td>59.9 ppb</td><td>62.1 ppb</td><td>76.0 ppb</td><td>65.1 ppb</td></tr>
<tr><td>Number of days</td><td>495</td><td>150</td><td>191</td><td>297</td></tr>
</table>

Table 2: Meteorological conditions, occurrence, and average ozone for the two nodes identified with high ozone concentrations for the Cook County monitors.

Conditions	Node 15	Node 16
Maximum temperature	> 77.62 °F	
Average afternoon temperature	> 85.27 °F	
Midday relative humidity	≤ 54.48%	
24-hour scalar wind run	≤ 727.29 km	> 727.29 km
Average ozone	66.1 ppb	58.3 ppb
Number of days	1352	572

The highest average ozone concentrations were observed for node 20 of the northern monitors. This node was characterized by maximum temperatures and average morning temperatures above 77.56 °F (7-10 am local time). Node 20 also had an average afternoon "v" wind vector that was greater than 2.41 meters per second (m/s) (i.e. from the south) as well as an average afternoon "u" vector that was less than or equal to -1.88 m/s (i.e. from the east). It contained 191 days of data and an average of 75.969 ppb in the data set under the aforementioned conditions. For Node 16, morning height values measured at 700 millibars altitude. Node 19 represented a maximum surface temperature above 85 °F. Node 21 represented morning temperatures above 77 °F (7-10 am local time), an average afternoon "v" wind vector greater than 2.41 meters per second (m/s) (i.e. from the south) as well as an average afternoon "u" vector equal to or greater than -1.88 m/s (i.e. from the east). It contained 297 days of data and an average of 65.088 ppb for ozone. Chart 1 below shows that for the four nodes identified among the northern monitors ozone values are trending lower for the most recent 15 years (2000-2014). Chart 2 shows the same trends for the Cook County monitors.

Chart 1

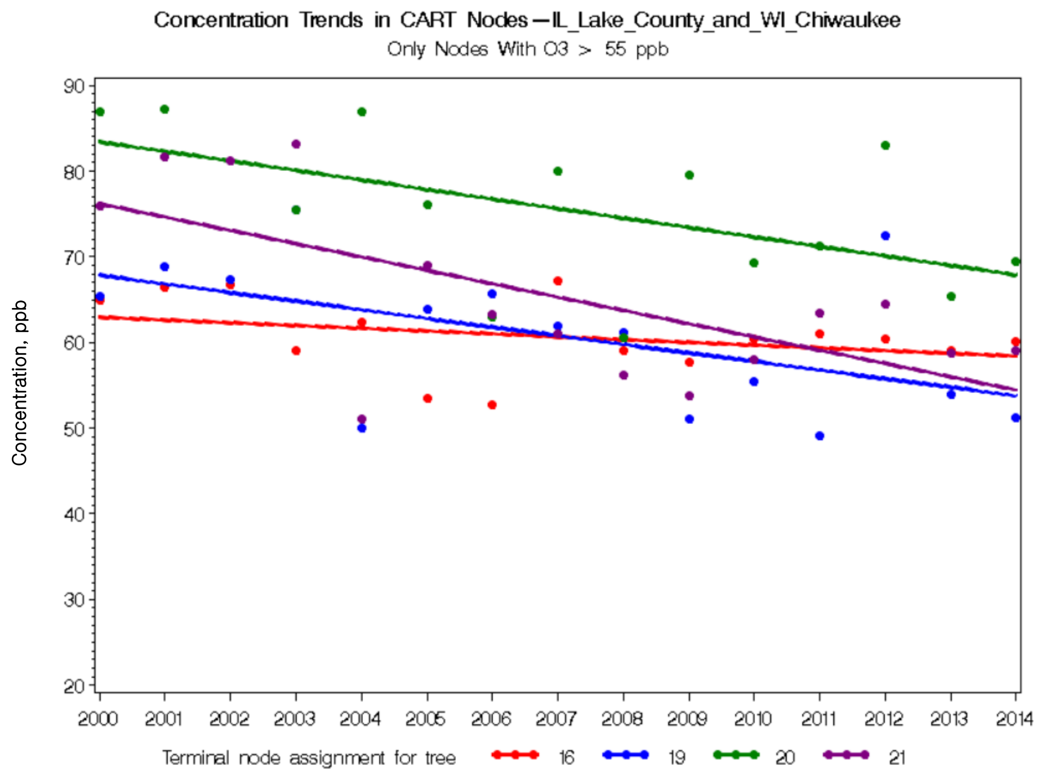
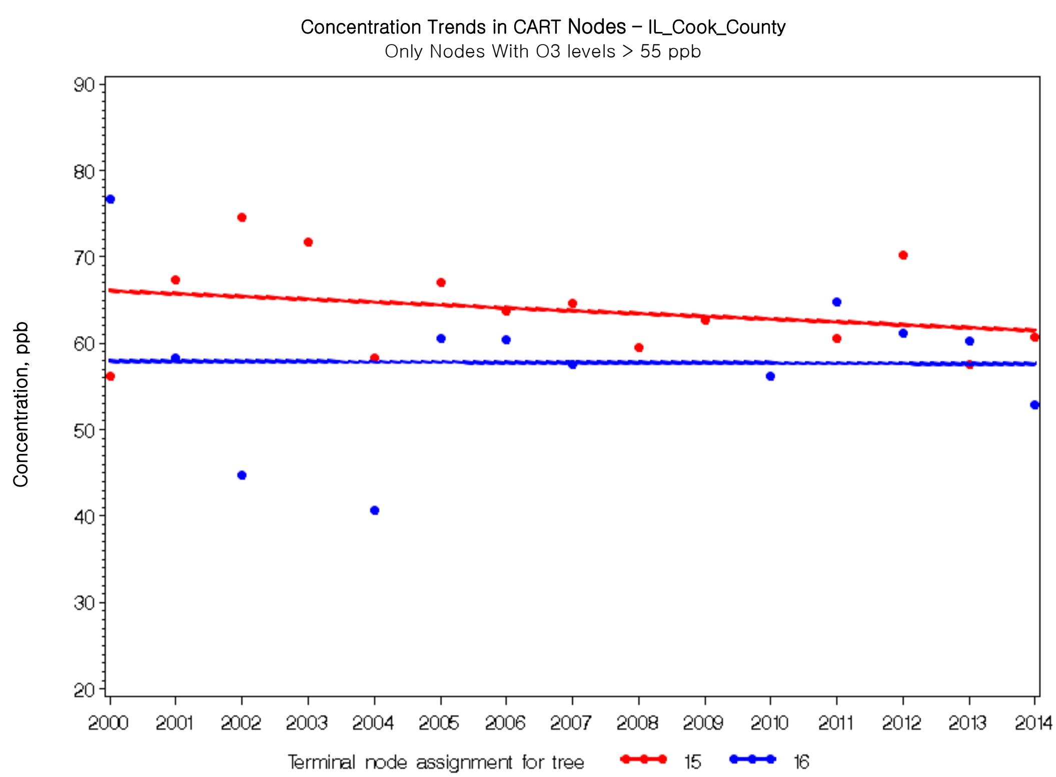


Chart 2



These analyses demonstrate that ozone concentrations, for a given set of high-ozone meteorological conditions, have decreased over time. In particular, this analysis shows that ozone concentrations have decreased on days with high average temperatures and the right combination of (mostly south-southeasterly) winds, low relative humidity and other characteristics. While maximum temperatures play an important role in the formation of ozone, the CART analysis reveals that other meteorological parameters (such as wind direction, wind speed, and morning temperature) also play significant roles in creating conditions conducive for ozone formation.

By using a CART analysis to analyze 8-hour ozone data in the Chicago-area, the influence of variations in meteorology can be mitigated such that comparisons of high ozone days with similar meteorological conditions can be made to determine if ozone values have decreased over time due to anthropogenic emission reductions. In general, ozone trends in the Chicago-area have declined. Furthermore, under meteorological conditions when monitored 8-hour ozone has historically been at its highest, ozone concentrations are lower under similar meteorological conditions. This analysis demonstrates that the observed reductions in ozone concentrations have not been driven solely by favorable meteorological conditions. These results further suggest that progress in reducing ozone precursor emissions is likely an important driver of the observed reductions in 8-hour ozone concentrations in the Chicago nonattainment area as well as areas to the north and east.

This page left intentionally blank.